نشریه پژوهشهای علوم دامی ایران
جلد 11، شماره 3، پاییز 1398، ص. 388-377

Iranian Journal of Animal Science Research
Vol. 11, No. 3, Fall 2019, p. 377-388

انتخاب ژنومی با استفاده از شبکه‌های عصبی مصنوعی بیزی و روش‌های پارامتری مطالعه مقایسه‌ای

سعید شادپور[1]، مجتبی طهمورث پور[2]*، محمد مهدی شریعتی[3]

تاریخ دریافت: 1397/05/08
تاریخ پذیرش: 1397/10/15

چکیده

در پژوهش حاضر توانایی پیش‌بینی ارزش های اصلاحی ژنومی شبکه‌های عصبی بیزی و روش‌های پارامتری در چهار معماری ژنتیکی شبیه سازی شده و چهار صفت واقعی موش با یکدیگر مقایسه شد. تعداد QTLها در معماری های ژنتیکی اول و سوم 50 و در معماری های ژنتیکی دوم و چهارم 500 در نظر گرفته شد. مقدار وراثت‌پذیری در معماری‌های ژنتیکی اول و دوم 0/3 و در معماری‌های ژنتیکی سوم و چهارم 0/7 بود. بیشترین صحت پیش‌بینی ژنومی حاصل از شبکه‌های عصبی بیزی در چهار معماری ژنتیکی شبیه‌سازی‌شده برابر با 0/644، 0/654، 0/800 و 0/81 بود. این مقادیر در روش‌های پارامتری برابر با 0/717، 0/685، 0/903 و 0/836 بود. حداکثر توانایی پیش‌بینی حاصل از شبکه‌های عصبی بیزی در پیش‌بینی وزن شش هفتگی، شیب رشد، شاخص توده بدنی و طول بدن به ترتیب برابر با 0/474، 0/359، 0/154 و 0/214 بود. توانایی پیش‌بینی روش‌های پارامتری در پیش‌بینی ژنومی این صفات مشابه و به‌طور متوسط برابر با 0/477، 0/369، 0/170 و 0/221 بود. میانگین مربعات خطای پیش‌بینی شبکه‌های عصبی بیزی در معماری‌های ژنتیکی شبیه‌سازی‌شده اندکی کمتر از روش‌های پارامتری و در داده‌های واقعی مشابه روش‌های پارامتری بود. مدت‌زمان اجرای شبکه‌های عصبی بیزی با افزایش تعداد نرون در لایه مخفی به‌صورت صعودی افزایش یافت. نتایج بدست آمده نشان داد با وجود بهتر بودن صحت و توانایی پیش بینی روش های پارامتری، شبکه‌های عصبی بیزی می توانند ارزش‌های اصلاحی ژنومی را با دقت مناسبی پیش‌بینی کنند. همچنین توانایی پیش‌بینی ژنومی شبکه‌های عصبی به صفات هدف، گونه موردنظر و معماری شبکه عصبی بستگی دارد.

واژه‌های کلیدی: ارزیابی ژنومی، روش‌های پارامتری، شبکه‌های عصبی، مقایسه کارایی.

مقدمه

انتخاب ژنومی با استفاده از نشانگرهای با تراکم زیاد، بخصوص SNPهایی که کل ژنوم را در برمی‌گیرند و اغلب در عدم تعادل پیوستگی با QTLهای مجاور خود هستند، ارزش ژنتیکی کل را پیش‌بینی می‌کند (17 و 27). در انتخاب ژنومی روش‌های گوناگونی برای برآورد ارزش‌های اصلاحی ژنومی معرفی و بکار گرفته‌شده‌اند. این روش‌ها در پیش‌بینی، فرضیات متفاوتی را در نظر می‌گیرند. روش‌های ناپارامتری ازجمله شبکه‌های عصبی مصنوعی نسبت به روش‌های پارامتری فرضیات کمتری در پیش‌بینی دارند و می‌توانند روابط غیرخطی را در پیش‌بینی‌های ژنومی اعمال نمایند. بنابراین ازنظر تئوری نسبت به تغییرات معماری‌های ژنتیکی مختلف پایدارتر بوده و توانایی ارائه پیش‌بینی‌های بهتری دارند (10) .

از عوامل مؤثر بر صحت پیش‌بینی‌ها در انتخاب ژنومی می‌توان سطح LD بین نشانگرها و QTLها، تعداد و نوع دام‌های جمعیت مرجع، وراثت‌پذیری صفت موردمطالعه، تعداد جایگاه‌های مؤثر بر صفت و توزیع اثرات آنها، قابلیت اعتماد اطلاعات فنوتیپی، روابط ژنتیکی دو جمعیت مرجع و اعتبارسنجی و مدل آنالیز را نام برد که نوع و میزان اثرات هر یک از این عوامل در مطالعات مختلف بررسی‌شده است (1 و 6 و 16 و 19).

در روش‌های پارامتری به‌کاررفته در انتخاب ژنومی، در رابطه با توزیع اثرات نشانگرها فرضیات مختلفی در نظر گرفته می‌شود. نتایج مطالعات پیشین نشان داده است که اعمال توزیع پیشین اثرات نشانگرها که با شکل توزیع اثرات واقعی نشانگرها همخوانی داشته باشد، می‌تواند صحت پیش‌بینی را تحت تأثیر قرار دهد چراکه پسین

1، 2 و 3- فارغ التحصیل دکتری، استاد و استادیار گروه علوم دامی دانشکده کشاورزی دانشگاه فردوسی مشهد
(* - نویسنده مسئول: tahmoores@um.ac.ir)
Doi:10.22067/ijasr.v11i3.74465

استنباط شده باید با صحت بالا اثرات واقعی نشانگرها در صفت موردنظر را تقریب بزند (2 و 16 و 30). درحالی‌که توزیع اثرات حقیقی نشانگرها ناشناخته است، انتخاب توزیع پیشینی که برای تقلید توزیع اثرات واقعی نشانگرها به‌اندازه کافی منعطف باشد، یک موضوع کلیدی در اجرای روش‌های پارامتری در انتخاب ژنومی باقی می‌ماند (30 و 41). همچنین برخی از فرضیات مدل‌های پارامتری از قبیل نرمال بودن، خطی بودن و استقلال متغیرهای توضیحی اغلب در پیش‌بینی ژنومی برقرار نیستند (15).

در سال‌های اخیر، برای پرداختن به نقاط ضعف روش‌های پارامتری روش‌های نیمه پارامتری ازجمله رگرسیون فضاهای هیلبرت با هسته بازآفرین[1] و برآوردگر نادارا- واتسون[2] (15) و روش‌های ناپارامتری ازجمله شبکه‌های عصبی مصنوعی[3] (14 و 18)، Boosting و جنگل تصادفی[4] (32) در انتخاب ژنومی معرفی‌شده‌اند. روش‌های ناپارامتری در رابطه با اثرات نشانگرها فرضیه‌ای را در نظر نمی‌گیرند، لذا ازنظر تئوری در مقابل تغییرات معماری ژنتیکی صفت پایدار بوده و به نظر می‌رسد می‌توانند ارزش اصلاحی ژنومی را نسبت به روش‌های پارامتری، با صحت بالاتری پیش‌بینی کنند.

باوجودآنکه مطالعات انتخاب ژنومی در حال حاضر دانش باارزشی در رابطه با تنوعات ژنی مؤثر بر صفات کمی را فراهم آورده‌اند، اما این مطالعات تنها بخشی از وراثت‌پذیری صفات پیچیده را توضیح می‌دهند (11). مسئله اساسی در توجیه وراثت‌پذیری گمشده می‌تواند عدم مطابقت معماری ژنتیکی صفات پیچیده با تکنیک‌های آماری به کار گرفته‌شده باشد (9 و 41). وجود اثرات غیر افزایشی و اثرات غیرخطی از درجات بالاتر می‌تواند یک روش تبیینی مهم در توجیه وراثت‌پذیری گمشده باشد. سازوکار ژنتیکی صفات پیچیده، ممکن است دربرگیرنده اثرات متقابل بین ژن‌ها، بین ژن‌ها و شرایط محیطی یا اثرات اپیستازی که به‌وسیله مدل‌های افزایشی به‌طور کامل اخذ نمی‌شوند، باشد (9). از این نظر تمایل فزاینده‌ای در استفاده از روش‌های نیمه پارامتری، ناپارامتری و یادگیری ماشین در پیش‌بینی ژنومی صفات کمی برای تخمین اثرات غیر افزایشی و همچنین اثرات متقابل ژنوتیپ و محیط وجود دارد (9 و 13 و 15).

هدف از پژوهش حاضر، بررسی کارایی شبکه‌های عصبی مصنوعی با الگوریتم آموزشی بیزی در پیش‌بینی ژنومی و مقایسه نتایج به‌دست‌آمده با روش‌های پارامتری بود.

1-Reproducing Kernel Hilbert Spaces Regression (RKHS)
2- Nadaraya-Watson Estimator (NWE)
3- Artificial Neural Networks (ANNs)
4- Random Forest (RF)

مواد و روش‌ها

در پژوهش حاضر، توانایی شبکه‌های عصبی مصنوعی با الگوریتم آموزش بیزی در پیش‌بینی ژنومی چهار معماری ژنتیکی شبیه‌سازی‌شده با اثرات افزایشی و 4 صفت واقعی موش شامل وزن شش هفتگی، شیب رشد، شاخص توده بدنی و طول بدن بررسی و نتایج به‌دست‌آمده با نتایج حاصل از روش‌های پارامتری GBLUP ، Bayes RR، Bayes A، Bayes B، Bayes C و Bayes L مقایسه شد.

در داده‌های شبیه‌سازی‌شده با تغییر مقدار وراثت‌پذیری (0/3 و 0/7) و تعداد QTLها (50 و 500) 4 معماری ژنتیکی مختلف شبیه‌سازی شد. مشخصات معماری‌های ژنتیکی شبیه‌سازی‌شده در جدول 1 نشان داده‌شده است.

در تمام این معماری‌های ژنتیکی توزیع اثرات SNPها، نرمال در نظر گرفته شد. برای شبیه‌سازی معماری‌های ژنتیکی با اثرات افزایشی از داده‌های ژنومی موش مارتینی و همکاران (26) استفاده شد. ابتدا با استفاده از نمونه‌برداری تصادفی 1000 نشانگر SNP از مجموعه نشانگرهای SNP تعیین ژنوتیپ شده در ژنوم موش جدا و سپس از بین این نشانگرهای نمونه‌برداری شده، به تعداد QTLهای در نظر گرفته‌شده در هر معماری ژنتیکی، نشانگر SNP با نمونه‌برداری مجدد انتخاب شدند. در ادامه به‌منظور شبیه‌سازی اثرات آللی QTLها، به‌اندازه 2 برابر تعداد QTLها در هر معماری ژنتیکی، اثرات آللی QTLها از توزیع نرمال با میانگین صفر و واریانس 0/01 نمونه‌برداری شدند. با در اختیار داشتن این اثرات، ارزش‌های اصلاحی واقعی اولیه حیوانات، متناظر با نوع و اثر آللی QTLهای مربوطه محاسبه شد. برای هر حیوان همچنین یک اثر خطای اولیه از توزیع نرمال با میانگین صفر و واریانس 0/01 نمونه‌برداری شد. در ادامه بردار ارزش‌های اصلاحی واقعی اولیه و اثرات خطای اولیه به ترتیب استاندارد و با توجه به مقدار توارث پذیری در نظر گرفته‌شده در هر معماری ژنتیکی، در انحراف معیار ژنتیکی افزایشی و اثرات خطا مناسب ضرب شدند. در پایان از جمع بردار ارزش‌های اصلاحی واقعی و اثرات خطا، بردار فنوتیپ حیوانات به دست آمد. پیش از انجام آنالیزها، 1000 نشانگر SNP نمونه‌برداری شده اولیه از مجموعه داده‌های ژنوتیپی حیوانات حذف شدند. محاسبات لازم برای شبیه‌سازی معماری‌های ژنتیکی مختلف با استفاده از کدهای نوشته‌شده در نرم‌افزار R انجام پذیرفت.

در بخش داده‌های واقعی از داده‌های ژنوم موش مارتینی و همکاران (26) شامل 1298 حیوان تعیین ژنوتیپ شده با 9265 نشانگر SNP استفاده شد. مارتینی و همکاران داده‌های یادشده را به شرح زیر ویرایش کردند.

جدول 1- تعداد QTL ها و مقدار وراثت‌پذیری در معماری‌های ژنتیکی شبیه‌سازی‌شده

Table 1- The values of heritability and QTL numbers in simulated genetic architectures

متغیر Variable	GA1 معماری ژنتیکی اول	GA2 معماری ژنتیکی دوم	GA3 معماری ژنتیکی سوم	GA4 معماری ژنتیکی چهارم
Number of QTLs تعداد QTL ها	50	500	50	500
Heritability وراثت‌پذیری	0.3	0.3	0.7	0.7

GA: Genetic Architecture

نقشه فیزیکی نشانگرهای SNP، با استفاده از بسته biomaRt نرم‌افزار R به آخرین نسخه ژنوم موش[1] بروز رسانی شد. در ادامه، تنها تنها از نشانگرهای SNP نقشه یابی شده بر روی GRCm38.p4 برای آنالیزها استفاده شد. در این داده‌های ژنوتیپی نرخ نشانگرهای گمشده برابر با 0/33 بود که با انتساب ژنوتیپی[2] جایگذاری شدند. پیش از آنالیز داده‌ها، نشانگرهای SNP با فراوانی آلل جزئی کمتر از 0/01 حذف شدند. کلیه مراحل انتساب ژنوتیپی، کدگذاری مجدد و تعیین کیفیت ژنوتیپ‌ها به‌طور همزمان و با استفاده از بسته Synbreed نرم‌افزار R انجام شد. برای پیش‌بینی ژنومی از رکوردهای فنوتیپی تصحیح‌شده برای اثرات جنس، فصل و ماه استفاده شد.

برای آموزش شبکه‌های عصبی بیزی از بسته brnn نرم‌افزار R استفاده شد. تعداد دوره ها[3] در شبکه‌های عصبی بیزی برابر با 30 در نظر گرفته شد و تعداد نرون‌ها در لایه مخفی از 1 تا 5 تغییر داده شد. تابع فعال‌ساز تانژانت هایپربولیک و تابع هدف میانگین مربعات خطای پیش‌بینی تصحیح‌شده بود. فرم ریاضی تابع تانژانت هایپربولیک در زیر نمایش داده‌شده است (10).

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} = \frac{e^{2x}-1}{e^{x}+1} = \frac{1-e^{-2x}}{1+e^{-2x}} \qquad (1)$$

Hyperbolic tangent:

هر شبکه عصبی بیزی شامل یک‌لایه ورودی، یک‌لایه مخفی و یک‌لایه خروجی بود. در لایه ورودی به ازای هر متغیر ورودی به شبکه عصبی یک وزن تصادفی اختصاص می‌یابد. وظیفه لایه ورودی، انتقال متغیرهای ورودی و وزن‌های اختصاص‌یافته آنها به نرون‌های لایه میانی است. در لایه میانی تعدادی نرون وجود دارد. هر نرون شبکه عصبی شامل سه بخش آکسون‌های ورودی، بدنه نرون و آکسون‌های خروجی است. وظیفه آکسون‌های ورودی، دریافت متغیرهای ورودی و وزن‌های مربوطه از لایه قبل و انتقال آنها به بدنه نرون است. در بدنه هر نرون، مجموع حاصل‌ضرب متغیرهای ورودی و وزن‌های متناظرشان با یک مقدار عرض از مبدأ (بایاس) جمع شده پس از عبور از تابع فعال‌ساز به خروجی نرون مربوطه تبدیل می‌شوند. به خروجی نرون‌های لایه میانی نیز یک وزن تصادفی اختصاص می‌یابد سپس این مقادیر به لایه خروجی انتقال می‌یابند. لایه خروجی متشکل از یک نرون با تابع فعال‌ساز خطی بود. تخمین تابع در شبکه‌های عصبی در طی فرآیند آموزش شبکه که شامل اصلاح وزن‌ها و بایاس‌های شبکه عصبی است، امکان‌پذیر می‌گردد. آموزش هر شبکه عصبی چندلایه در دو مرحله انتشار روبه‌جلو و پس انتشار رو به عقب از طریق الگوریتم آموزش شبکه انجام می‌شود. در مرحله انتشار روبه‌جلو، ورودی‌ها و وزن‌های اختصاص‌یافته به هر یک از آنها، از یک لایه به لایه بعد از لایه ورودی به سمت لایه خروجی منتشر می‌شود. پس‌ازاین مرحله، خروجی شبکه با خروجی ایده آل مقایسه شده و میزان خطای شبکه محاسبه می‌گردد. در مرحله انتشار رو به عقب، مقدار خطای شبکه رو به عقب (از سمت خروجی شبکه عصبی به سمت ورودی) انتشار می‌یابد. پس از انتشار رو به عقب، الگوریتم آموزش شبکه عصبی بااطلاع از گرادیان تابع هزینه نسبت به وزن‌ها و بایاس‌های شبکه عصبی، این مقادیر را به‌روزرسانی می‌نماید. این فرآیند تا پایان شرط توقف آموزش شبکه عصبی که یکی از شروط رسیدن به مقدار آستانه خطا قابل‌قبول، حداکثر تکرار (تعداد مراحل بروز رسانی وزن‌ها و بایاس‌ها) و یا حداکثر مدت‌زمان اجرا است، ادامه می‌یابد (25).

در شبکه های عصبی بیزی تابع هدف تصحیح شده به شکل زیر است.

$$F - \beta E_D(D|w,M) + \propto E_w(w|M) \qquad (2)$$

در رابطه بالا $E_D(D|w,M)$ مجموع مربعات باقی مانده و $E_w(w|M)$ مجموع مربعات وزن های شبکه است. M و w به ترتیب معماری و وزن های شبکه عصبی و α و β پارامترهای تنظیم شبکه هستند که تعادل نیکویی برازش و میزان پیچیدگی مدل برازش یافته را کنترل می نمایند. در چهارچوب بیز تجربی، وزن های بهینه مقادیری هستند که چگالی پسین شرطی $P = (w|D,\propto,\beta,M)$ را بیشینه نمایند که معادل با کمینه سازی تابع هدف F خواهد بود. کمینه سازی F برابر با یافتن بیشینه

1- Mus musculus, assembly GRCm38.p4
2- Imputation
3- Epoch

برآوردهای پسین w (w^{MP}) است که E_D را با استفاده از الگوریتم انتشار به عقب کمینه می‌نماید (26). تئوری بیز چگالی پسین α و β را بصورت $P(\alpha,\beta|D,M)=\frac{P(D|\alpha,\beta,M)P(\alpha,\beta|M)}{P(D|M)}$ نتیجه می‌دهد.

اگر چگالی پیشین $P(\alpha,\beta|M)$ یکنواخت در نظر گرفته شود، بیشینه‌سازی $P(\alpha,\beta|D,M)$ نسبت به α و β معادل بیشینه‌سازی $P(D|\alpha,\beta,M)$ خواهد بود. بهینه‌سازی پارامترهای تنظیم، نیاز به ماتریس هیسین تابع F تخمین زده شده در نقطه W^{MP} دارد (40). اما همانطور که توسط مک کی (25) پیشنهاد شده است اگر از الگوریتم آموزش لونبرگ- مارکوارت برای یافتن کمینه F استفاده شده باشد، محاسبه مستقیم ماتریس هیسین ضروری نیست و می‌توان از تخمین گوس- نیوتن ماتریس هیسین استفاده نمود (22 و 34 و 37). به طور خلاصه ماتریس هیسن بصورت $H=JJ'$ تقریب زده می‌شود. J ماتریس ژاکوبین است که شامل مشتقات اول خطاهای شبکه نسبت به پارامترهای شبکه (وزن‌ها و بایاس‌ها) می باشد. پارامترهای شبکه به صورت زیر بروزرسانی می شوند.

$$w_{k+1}=w_k-(JJ'+\mu I)\times J'e \qquad (٣)$$

در رابطه بالا μ فاکتور تعدیل لونبرگ است که در هر تکرار تنظیم و فرآیند بهینه‌سازی را هدایت می‌نماید.

در شبکه‌های عصبی بیزی مقادیر بهینه پارامترهای تنظیم α و β به‌صورت زیر محاسبه می‌شوند.

$$\alpha^{MP}=\frac{\gamma}{2E_W(w^{MP})}\ ,\quad \beta^{MP}=\frac{n-\gamma}{2E_D(w^{MP})} \qquad (4)$$

γ تعداد پارامترهای مؤثر در شبکه عصبی است که در رابطه زیر صدق می‌نماید.

$$0\le\gamma=m-2\alpha^{MP}tr(H^{MP})^{-1}\le m \qquad (5)$$

در رابطه بالا H^{MP} بیشینه برآورد پسین و m تعداد کل پارامترهای شبکه عصبی است.

از ماتریس روابط خویشاوندی ژنومی حیوانات به‌عنوان ورودی شبکه‌های عصبی استفاده شد که بر اساس رابطه زیر به دست آمد.

$$G=\frac{ZZ'}{P} \qquad (6)$$

در رابطه بالا Z ماتریس استانداردشده نشانگرهای SNP و P برابر با تعداد کل نشانگرهای SNP است.

هر یک از عناصر ماتریس Z بر اساس رابطه زیر محاسبه شد.

$$Z_{ij}=\frac{X_{ij}-\mu_{.j}}{\sigma_{X_{.j}}} \qquad (7)$$

در رابطه بالا X_{ij} نشان‌دهنده ژنوتیپ حیوان i ام در نشانگر SNP j ام و $\mu_{X_{.j}}$ و $\sigma_{X_{.j}}$ به ترتیب نشان‌دهنده میانگین و انحراف معیار نشانگر SNP j ام است.

برای اجرای روش‌های پارامتری، بسته آماری BGLR به کار گرفته شد (8). در تمام این روش‌ها تعداد تکرارها برابر با 12000 در نظر گرفته شد. از این تعداد تکرار، 2000 تکرار نخست به‌عنوان دوره گرم شدگی حذف شدند. برای اطمینان از همگرایی نمونه‌های گیبس از رسم پلات و خودهمبستگی بین نمونه‌ها استفاده شد. جزئیات روش‌های پارامتری مورداستفاده در (35) ارائه‌شده است. جهت مقایسه کارایی شبکه‌های عصبی بیزی و روش‌های پارامتری در پیش‌بینی ژنومی معماری‌های ژنتیکی شبیه‌سازی‌شده و صفات واقعی مورد بررسی از معیارهای صحت پیش‌بینی یا توانایی پیش‌بینی، میانگین مربعات خطای پیش‌بینی و مدت‌زمان اجرا برحسب ثانیه استفاده شد. صحت پیش‌بینی درداده های شبیه‌سازی‌شده برابر با ضریب همبستگی پیرسون ارزش‌های ژنومی پیش‌بینی‌شده و مقادیر ارزش‌های اصلاحی واقعی بود. درداده های واقعی ضریب همبستگی بین ارزش‌های ژنومی پیش‌بینی‌شده و مقادیر فنوتیپی مشاهده‌شده توانایی پیش‌بینی نامیده شد. در پیش‌بینی ژنومی هر یک از معماری‌های ژنتیکی شبیه‌سازی‌شده و صفات واقعی در پژوهش حاضر ابتدا داده‌های موجود با استفاده از نمونه‌برداری تصادفی به 10 مجموعه آموزش و آزمون به ترتیب متشکل از 1000 و 296 داده تقسیم شدند و ارزیابی مدل‌های بررسی‌شده بر اساس عملکردشان در پیش‌بینی این 10 مجموعه داده (تکرار) انجام پذیرفت.

نتایج و بحث

صحت پیش‌بینی شبکه‌های عصبی بیزی و روش‌های پارامتری در معماری‌های ژنتیکی شبیه‌سازی‌شده با اثرات افزایشی در جدول 2 نشان داده‌شده است. بیشترین صحت پیش‌بینی حاصل از شبکه‌های عصبی بیزی در 4 معماری ژنتیکی شبیه‌سازی‌شده به ترتیب برابر با 0/644 (4 نرون در لایه مخفی)، 0/654 (4 نرون در لایه مخفی)، 0/800 (1 نرون در لایه مخفی) و 0/81 (ا نرون در لایه مخفی) بود. این مقادیر در روش‌های پارامتری به ترتیب برابر با 0/717 (Bayes B)، 0/685 (Bayes A)، 0/903 (Bayes B) و 0/836 (Bayes B) بود. این نتایج، نشان‌دهنده برتری روش‌های پارامتری نسبت به شبکه‌های عصبی بیزی ازنظر صحت پیش‌بینی در معماری‌های ژنتیکی با اثرات افزایشی بررسی‌شده است. این یافته با نتایج محمدی و همکاران (29) که صحت پیش‌بینی در صفات تولید شیر،

تولید چربی، تولید پروتئین و تعداد سلول‌های سوماتیک در جمعیت گاوهای هلشتاین تعیین ژنوتیپ شده متشکل از 345 گاو نر را بررسی کردند، همخوانی دارد. هووارد و همکاران (21) در یک مطالعه مقایسه‌ای صحت پیش‌بینی 4 روش ناپارامتری شبکه‌های عصبی مصنوعی، رگرسیون بردار پشتیبان، رگرسیون فضاهای هیلبرت با هسته بازآفرین و برآوردگر نادارایا- واتسون را با روش‌های پارامتری مقایسه کردند. نتایج مطالعه آنها نیز نشان‌دهنده برتری روش‌های پارامتری نسبت به روش‌های ناپارامتری در معماری‌های ژنتیکی افزایشی بود. در معماری‌های ژنتیکی افزایشی تأثیر آللی تنوعات ژنی مؤثر بر صفات مستقل از یکدیگرند. در مدل‌های پارامتری نیز این اثرات مستقل فرض می‌شوند بنابراین، می‌توان انتظار داشت در معماری‌های ژنتیکی افزایشی، روش‌های پارامتری نسبت به روش‌های ناپارامتری پیش‌بینی‌های بهتری ارائه دهند. نتایج به‌دست‌آمده، همچنین نشان‌دهنده برتری روش‌های پارامتری جریمه‌ای همراه با انتخاب متغییر به‌ویژه Bayes B در معماری‌های ژنتیکی با تعداد کم QTL (معماری اول و سوم) نسبت به روش‌های صرفاً جریمه‌ای مانند GBLUP ازنظر صحت پیش‌بینی است. این یافته‌ها در مطالعات پیشین نیز گزارش‌شده‌اند. (3 و 7 و 28). نتایج مطالعات پیشین نشان داده است که در پیش‌بینی برخی از صفات واقعی که تحت تأثیر ژن‌های بزرگ اثر هستند، روش‌های جریمه‌ای همراه با انتخاب متغییر نسبت به روش‌های صرفاً جریمه‌ای عملکرد بهتری دارند ازجمله این صفات می‌توان به چربی شیر اشاره کرد که تحت تأثیر ژن بزرگ اثر DGAT1 است (23). با مقایسه صحت پیش‌بینی حاصل از روش‌های مختلف در معماری‌های ژنتیکی (اول و سوم) و (دوم و چهارم) می‌توان دریافت که با افزایش مقدار وراثت‌پذیری در معماری‌های ژنتیکی با تعداد QTL برابر، صحت پیش‌بینی تمامی روش‌ها افزایش‌یافته است. این نتیجه در تطابق با یافته‌های مطالعات پیشین است (12 و 24 و 31). وراثت‌پذیری یک صفت درواقع نشان‌دهنده نسبت اندازه سیگنال ارزش‌های افزایشی به‌اندازه سیگنال خطا است. به‌عبارت‌دیگر، بیشتر بودن وراثت‌پذیری به معنای کمتر بودن اندازه سیگنال خطای داده‌های ورودی به مدل است که منجر به بهبود صحت پیش‌بینی روش‌های مختلف می‌گردد.

توانایی پیش‌بینی شبکه‌های عصبی بیزی و روش‌های پارامتری در پیش‌بینی داده‌های واقعی در جدول 3 ارائه‌شده است. حداکثر توانایی پیش‌بینی حاصل از شبکه‌های عصبی بیزی در پیش‌بینی وزن شش هفتگی، شیب رشد، شاخص توده بدنی و طول بدن به ترتیب برابر با 0/474 (1 نرون در لایه مخفی)، 0/359 (4 نرون در لایه مخفی)، 0/154 (1 نرون در لایه مخفی)، 0/214 (4 نرون در لایه مخفی) بود. توانایی پیش‌بینی روش‌های پارامتری در پیش‌بینی ژنومی این صفات مشابه بود. این یافته با نتایج پژوهش‌های پیشین که گزارش کردند در پیش‌بینی ژنومی بیشتر صفات واقعی روش‌های جریمه‌ای مشابه و یا حتی بهتر از روش‌های جریمه‌ای توأم با انتخاب متغییر عمل می‌نمایید، همخوانی دارد. دلیل این امر معماری پلی ژنیک این صفات و توزیع نرمال تأثیرات ژنی است (4 و 23 و 39). توانایی پیش‌بینی روش‌های پارامتری و شبکه‌های عصبی بیزی در پیش‌بینی این صفات نیز مشابه بود به‌طوری‌که تفاوت بهترین توانایی پیش‌بینی حاصل از شبکه‌های عصبی بیزی و روش‌های پارامتری در سه صفت وزن شش هفتگی، شیب رشد و طول بدن کمتر از 1 ٪ بود. این مقدار برای صفت شاخص توده بدنی اندکی بیشتر و برابر با 1/8% بود. در رابطه با توانایی پیش‌بینی ژنومی شبکه‌های عصبی، مطالعات مختلفی انجام‌شده است.

جدول 2- صحت پیش‌بینی شبکه‌های عصبی بیزی و روش‌های معیار در معماری‌های ژنتیکی شبیه‌سازی‌شده

Table 2- The prediction accuracy of Bayesian Regularized Neural Networks and benchmark methods on the simulated data

روش‌های مورد مطالعه Study methods		معماری ژنتیکی اول A1	معماری ژنتیکی دوم A2	معماری ژنتیکی سوم A3	معماری ژنتیکی چهارم A4
شبکه عصبی بیزی BRNN	NN=1	0.641 (0.0155)	0.638 (0.0240)	0.800 (0.0117)	0.810 (0.0111)
	NN=2	0.641 (0.0224)	0.643 (0.0251)	0.784 (0.0168)	0.788 (0.0204)
	NN=3	0.64 (0.0233)	0.647 (0.0243)	0.740 (0.0269)	0.744 (0.0169)
	NN=4	0.644 (0.0208)	0.654 (0.0255)	0.761 (0.0206)	0.764 (0.0237)
	NN=5	0.633 (0.0252)	0.653 (0.0241)	0.740 (0.0210)	0.736 (0.0228)
روش‌های پارامتری PM	GBLUP	0.683 (0.0200)	0.684 (0.0213)	0.820 (0.0113)	0.830 (0.0103)
	Bayes RR	0.682 (0.0198)	0.683 (0.0211)	0.819 (0.0115)	0.830 (0.0102)
	Bayes A	0.694 (0.0199)	0.685 (0.0210)	0.859 (0.0096)	0.833 (0.0102)
	Bayes B	0.717 (0.0205)	0.684 (0.0213)	0.903 (0.0090)	0.836 (0.0100)
	Bayes C	0.688 (0.0190)	0.683 (0.0213)	0.888 (0.0106)	0.831 (0.0105)
	Bayes L	0.687 (0.0201)	0.684 (0.0207)	0.834 (0.0102)	0.832 (0.0103)

BRNN: Bayesian Regularized Neural Networks
PM: Parametric Methods
NN: Number of neurons in hidden layer

جدول 3- توانایی پیش‌بینی شبکه‌های عصبی بیزی و روش‌های معیار در داده‌های موش

Table 4- The predictive ability of Bayesian Regularized Neural Networks and benchmark methods in genomic enabled prediction on the mouse data

روش‌های مورد مطالعه Study methods		وزن شش هفتگی W6W	شیب رشد Growth Slope	شاخص توده بدنی BMI	طول بدن Body Length
شبکه عصبی بیزی BRNN	NN=1	0.474 (0.0171)	0.339 (0.0251)	0.154 (0.0154)	0.211 (0.0169)
	NN=2	0.468 (0.0204)	0.356 (0.0252)	0.149 (0.0130)	0.201 (0.0228)
	NN=3	0.438 (0.0179)	0.358 (0.0212)	0.145 (0.0140)	0.207 (0.0228)
	NN=4	0.464 (0.0184)	0.359 (0.0209)	0.144 (0.0134)	0.214 (0.0223)
	NN=5	0.462 (0.0198)	0.355 (0.0204)	0.144 (0.0147)	0.206 (0.0228)
روش‌های پارامتری PM	GBLUP	0.477 (0.0190)	0.368 (0.0224)	0.170 (0.0137)	0.222 (0.0167)
	Bayes RR	0.477 (0.0191)	0.368 (0.0221)	0.170 (0.0141)	0.220 (0.0159)
	Bayes A	0.477 (0.0189)	0.369 (0.0222)	0.172 (0.0142)	0.221 (0.0165)
	Bayes B	0.476 (0.0183)	0.370 (0.0219)	0.164 (0.0142)	0.222 (0.0163)
	Bayes C	0.477 (0.0188)	0.368 (0.0220)	0.170 (0.0136)	0.220 (0.0163)
	Bayes L	0.476 (0.0187)	0.368 (0.0216)	0.171 (0.0130)	0.218 (0.0175)

BRNN: Bayesian Regularized Neural Networks
PM: Parametric Methods
NN: Number of neurons in hidden layer

هس لات و همکاران (20) توانایی پیش‌بینی شبکه‌های عصبی مصنوعی و روش‌های پارامتری را در سه مجموعه داده جو، ذرت و گندم بررسی کردند. نتایج به‌دست‌آمده نشان داد در بیشتر ارزیابی‌های انجام‌شده در داده‌های گندم، شبکه‌های عصبی مصنوعی نسبت به روش‌های پارامتری توانایی پیش‌بینی بهتری داشتند اما در ارزیابی‌های انجام‌شده بر روی داده‌های جو و ذرت برتری با روش‌های پارامتری بود.

بهتر بودن توانایی پیش‌بینی ژنومی شبکه‌های عصبی مصنوعی نسبت به روش‌های پارامتری در بیشتر ارزیابی‌های ژنومی انجام‌شده بر روی داده‌های گندم توسط پرز - رودریگز و همکاران (36) و کروسا و همکاران (5) نیز گزارش‌شده است. تاسل و همکاران (38) توانایی پیش‌بینی شبکه‌های عصبی بیزی و روش‌های پارامتری در پیش‌بینی تعداد نتاج در هر زایش در دو لاین خالص و یک لاین آمیخته خوک را بررسی و گزارش کردند توانایی پیش‌بینی شبکه‌های عصبی بیزی در پیش‌بینی تعداد نتاج در هر زایش در لاین‌های خالص کمتر و یا مشابه روش‌های پارامتری بود اما در لاین آمیخته شبکه‌های عصبی بیزی توانایی پیش‌بینی بهتری داشتند. اوکات و همکاران (33) توانایی شبکه‌های عصبی بیزی و روش Bayes cπc (Bayes cπ با π برابر با صفر) در پیش‌بینی تفاوت نتاج مورد انتظار نمره ماربلینگ گاوهای آنگوس در دو مجموعه داده 3k و 700 نشانگر SNP انتخاب‌شده را مقایسه کردند. نتایج نشان داد، در داده‌های 3k عملکرد شبکه‌های عصبی بیزی تنها اندکی بهتر از Bayes cπc بود و برعکس در داده‌های 700 نشانگر SNP انتخاب‌شده عملکرد Bayes cπc اندکی بهتر از شبکه‌های عصبی بیزی بود. آنها دلیل بهتر نبودن توانایی شبکه‌های عصبی نسبت به روش Bayes cπc را رابطه خطی تفاوت نتاج مورد انتظار نمره ماربلینگ و اثرات نشانگرهای SNP گزارش کردند. اریت و همکاران (10) با بررسی توانایی شبکه‌های عصبی پرسپترون در پیش‌بینی ژنومی صفات تولید شیر، چربی و پروتئین گزارش کردند که در صورت استفاده از ماتریس روابط خویشاوندی ژنومی و یا نمرات اجزای اصلی به‌عنوان ورودی شبکه‌های عصبی توانایی پیش‌بینی شبکه‌های عصبی با تعداد نرون متفاوت، تفاوت اندکی داشتند. بااین‌وجود، در شبکه‌های عصبی غیرخطی افزایش تعداد نرون در لایه مخفی منجر به اندکی بهبود در توانایی پیش‌بینی گردید. در بررسی توانایی شبکه‌های عصبی بیزی در پیش‌بینی ژنومی صفات تولید شیر، تولید چربی و تولید پروتئین گاوهای جرسی و صفت عملکرد دانه گندم در مطالعه جیانولا و همکاران (14)، شبکه‌های عصبی غیرخطی با بیش از یک نرون در لایه مخفی نسبت به شبکه‌های عصبی خطی توانایی پیش‌بینی بهتری داشتند. این ویژگی در پیش‌بینی صفت عملکرد دانه گندم بارزتر بود. نتایج پژوهش حاضر و مطالعات ذکرشده پیشین، نشان می‌دهد توانایی پیش‌بینی شبکه‌های عصبی در مقایسه با روش‌های پارامتری به صفات هدف و گونه موردنظر بستگی داشته و به نظر می‌رسد شبکه‌های عصبی نسبت به روش‌های پارامتری در پیش‌بینی ژنومی صفات در نژادهای آمیخته، عملکرد بهتری دارند.

میانگین مربعات خطای پیش‌بینی شبکه‌های عصبی بیزی و روش‌های پارامتری در معماری‌های ژنتیکی شبیه‌سازی‌شده با اثرات افزایشی در جدول 4 ارائه‌شده است. حداقل میانگین مربعات خطای پیش‌بینی حاصل از شبکه‌های عصبی بیزی در 4 معماری ژنتیکی شبیه‌سازی‌شده به ترتیب برابر با 0/185 (5 نرون در لایه مخفی)، 0/167 (4 و 5 نرون در لایه مخفی)، 0/287 (2 نرون در لایه مخفی) و 0/289 (1 نرون در لایه مخفی) بود. حداقل مربعات خطای پیش‌بینی حاصل از روش‌های پارامتری در معماری‌های ژنتیکی

شبیه‌سازی‌شده به ترتیب 0/178 (Bayes B)، 0/187 (GBLUP, Bayes RR, Bayes A, Bayes B)، 0/437 (GBLUP, Bayes RR) و 0/313 (Bayes B) بود. این نتایج نشان‌دهنده برتری نسبی شبکه‌های عصبی بیزی نسبت به روش‌های پارامتری ازنظر میانگین مربعات خطای پیش‌بینی به‌عنوان شاخصی از برازش کل است.

میانگین مربعات خطای پیش‌بینی شاخص مناسبی برای ارزیابی کارایی روش‌های مختلف در برازش مدل است چراکه همزمان دربردارنده میزان صحت و اریبی مدل برازش یافته می‌باشد. با در نظر گرفتن نتایج جداول 2 و 4 می‌توان نتیجه گرفت که مدل‌های برازش یافته با شبکه عصبی بیزی نسبت روش‌های پارامتری صحت کمتری داشته اما نااریب تر هستند. این مطلب می‌تواند به دلیل اعمال جریمه بیشتر در روش‌های پارامتری نسبت به شبکه‌های عصبی بیزی باشد که منجر به بیشتر شدن میانگین مربعات خطای پیش‌بینی می‌گردد.

میانگین مربعات خطای پیش‌بینی شبکه‌های عصبی بیزی و روش‌های پارامتری در پیش‌بینی ژنومی داده‌های موش در جدول 5 آورده شده است. طبق نتایج حاضر، عملکرد شبکه‌های عصبی بیزی و روش‌های پارامتری ازنظر میانگین مربعات خطای پیش‌بینی در پیش‌بینی صفات شیب رشد و شاخص توده بدنی مشابه بود. در دو صفت وزن شش هفتگی و طول بدن نیز تفاوت عملکرد شبکه‌های عصبی بیزی و روش‌های پارامتری ازنظر میانگین مربعات خطای پیش‌بینی بسیار ناچیز بود. این موضوع بیانگر تشابه کارایی شبکه‌های عصبی بیزی و روش‌های پارامتری در پیش‌بینی صفات مورد بررسی ازنظر میانگین مربعات خطای پیش‌بینی بود.

جدول 4- میانگین مربعات خطای پیش‌بینی شبکه‌های عصبی بیزی و روش‌های معیار در معماری‌های ژنتیکی شبیه‌سازی‌شده

Table 3- The mean squared error of prediction of Bayesian Regularized Neural Networks and benchmark methods on the simulated data.

روش‌های مورد مطالعه Study methods		معماری ژنتیکی اول A1	معماری ژنتیکی دوم A2	معماری ژنتیکی سوم A3	معماری ژنتیکی چهارم A4
شبکه عصبی بیزی BRNN	NN=1	0.243 (0.0175)	0.227 (0.0110)	0.288 (0.0180)	0.289 (0.0168)
	NN=2	0.204 (0.0111)	0.188 (0.0112)	0.287 (0.0155)	0.290 (0.0240)
	NN=3	0.190 (0.0077)	0.178 (0.0073)	0.321 (0.0323)	0.316 (0.0213)
	NN=4	0.183 (0.0078)	0.167 (0.0057)	0.295 (0.0235)	0.292 (0.0254)
	NN=5	0.185 (0.0091)	0.167 (0.0054)	0.319 (0.0262)	0.322 (0.0271)
روش‌های پارامتری PM	GBLUP	0.195 (0.0072)	0.187 (0.0091)	0.437 (0.0615)	0.325 (0.0128)
	Bayes RR	0.195 (0.0069)	0.187 (0.0087)	0.437 (0.0623)	0.327 (0.0129)
	Bayes A	0.187 (0.0084)	0.187 (0.0081)	0.543 (0.1214)	0.319 (0.0136)
	Bayes B	0.178 (0.0103)	0.187 (0.0093)	0.678 (0.2078)	0.313 (0.0141)
	Bayes C	0.194 (0.0070)	0.188 (0.0092)	0.753 (0.2632)	0.319 (0.0119)
	Bayes L	0.195 (0.0069)	0.188 (0.0074)	0.443 (0.0719)	0.324 (0.0128)

BRNN: Bayesian Regularized Neural Networks
PM: Parametric Methods
NN: Number of neurons in hidden layer

جدول 5- میانگین مربعات خطای پیش‌بینی شبکه‌های عصبی بیزی و روش‌های معیار در داده‌های موش

Table 5- The mean squared error of prediction of Bayesian Regularized Neural Networks and benchmark methods in genomic enabled prediction on the mouse data

روش‌های مورد مطالعه Study methods		وزن شش هفتگی W6W	شیب رشد Growth Slope	شاخص توده بدنی BMI	طول بدن Body Length
شبکه عصبی بیزی BRNN	NN=1	0.004 (0.0001)	0.002 (0.0001)	0.003 (0.0001)	0.227 (0.0076)
	NN=2	0.006 (0.0006)	0.002 (0.0000)	0.003 (0.0001)	0.215 (0.0077)
	NN=3	0.008 (0.0005)	0.002 (0.0000)	0.003 (0.0001)	0.214 (0.0077)
	NN=4	0.007 (0.0003)	0.002 (0.0000)	0.003 (0.0001)	0.214 (0.0079)
	NN=5	0.007 (0.0002)	0.002 (0.0000)	0.003 (0.0001)	0.213 (0.0078)
روش‌های پارامتری PM	GBLUP	0.005 (0.0001)	0.002 (0.0000)	0.003 (0.0001)	0.216 (0.0064)
	Bayes RR	0.005 (0.0001)	0.002 (0.0000)	0.003 (0.0001)	0.216 (0.0063)
	Bayes A	0.005 (0.0001)	0.002 (0.0000)	0.003 (0.0001)	0.216 (0.0063)
	Bayes B	0.005 (0.0001)	0.002 (0.0000)	0.003 (0.0001)	0.215 (0.0063)
	Bayes C	0.005 (0.0001)	0.002 (0.0000)	0.003 (0.0001)	0.215 (0.0062)
	Bayes L	0.005 (0.0001)	0.002 (0.0000)	0.003 (0.0001)	0.214 (0.0065)

BRNN: Bayesian Regularized Neural Networks
PM: Parametric Methods
NN: Number of neurons in hidden layer

میانگین مدت‌زمان اجرای شبکه‌های عصبی بیزی و روش‌های معیار در داده‌های شبیه‌سازی‌شده و صفات واقعی به ترتیب در جداول 6 و 7 گزارش‌شده است. طبق نتایج به‌دست‌آمده، در هر دو مجموعه معماری‌های ژنتیکی شبیه‌سازی‌شده و صفات واقعی، زمان محاسبات در شبکه عصبی بیزی با افزایش تعداد نرون به‌صورت تصاعدی افزایش یافت. مدت‌زمان اجرای شبکه‌های عصبی بیزی با یک نرون در لایه مخفی کمتر از روش‌های پارامتری بود. میانگین زمان محاسبات روش‌های پارامتری به‌استثنای GBLUP مشابه بود و روش GBLUP به علت معکوس کردن ماتریس خویشاوندی ژنومی و به دست آوردن اثرات مارکرها از ارزش‌های اصلاحی ژنومی زمان محاسباتی بیشتری را به خود اختصاص داد. مدت‌زمان اجرای شبکه‌های عصبی بیزی با 2 نرون در لایه مخفی همچنان کمتر از روش GBLUP بود. این نتایج نشان‌دهنده قابل‌قبول بودن هزینه محاسباتی شبکه‌های عصبی بیزی با 1 تا 2 نرون در لایه مخفی در پیش‌بینی ژنومی است. پیش‌بینی ژنومی با استفاده از شبکه‌های عصبی بیزی با تعداد نرون بیشتر حقیقتاً چالش‌برانگیز است و بهبود کارایی آنها ازنظر هزینه محاسباتی قبل از استفاده عملی در انتخاب ژنومی، ضروری خواهد بود. اوکات و همکاران (34) نیز هزینه محاسباتی بالای شبکه‌های عصبی بیزی در پیش‌بینی ژنومی را گزارش کردند.

جدول 6- میانگین مدت‌زمان اجرای شبکه‌های عصبی بیزی و روش‌های معیار در داده‌های شبیه‌سازی‌شده (برحسب ثانیه)

Table 6- The average computation time of Bayesian Regularized Neural Networks and benchmark methods on the simulated data (per second)

روش‌های مورد مطالعه Study methods		معماری ژنتیکی اول A1	معماری ژنتیکی دوم A2	معماری ژنتیکی سوم A3	معماری ژنتیکی چهارم A4
شبکه عصبی بیزی BRNN	NN=1	37.32 (1.22)	37.23 (1.33)	38.36 (1)	39.12 (0.44)
	NN=2	280.57 (0.38)	278.25 (1.77)	278.17 (1.27)	280 (1.18)
	NN=3	892.75 (2.02)	885 (5.67)	881.55 (5.26)	888.37 (5.01)
	NN=4	2040.4 (4.55)	2019.13 (16.58)	2004.24 (12.57)	2026.37 (9.78)
	NN=5	3877.41 (5.87)	3802.36 (31.66)	3812.98 (24.57)	3834.07 (23.06)
روش‌های پارامتری PM	GBLUP	302.61 (0.64)	301.93 (0.16)	302.03 (0.17)	302.29 (0.18)
	Bayes RR	125.03 (0.41)	124.88 (0.06)	124.83 (0.04)	124.82 (0.03)
	Bayes A	136.16 (0.4)	135.93 (0.06)	135.81 (0.07)	135.76 (0.05)
	Bayes B	138.07 (3.2)	145.82 (2.85)	118.17 (0.77)	151.43 (2.45)
	Bayes C	142.78 (1.59)	143.14 (0.87)	104.04 (3.02)	143.77 (0.99)
	Bayes L	141.17 (0.44)	140.98 (0.08)	140.92 (0.13)	140.97 (0.05)

BRNN: Bayesian Regularized Neural Networks
PM: Parametric Methods
NN: Number of neurons in hidden layer

جدول 7- میانگین مدت‌زمان اجرای شبکه‌های عصبی بیزی و روش‌های معیار در داده‌های موش (برحسب ثانیه)

Table 7- The average computation time of Bayesian Regularized Neural Networks and benchmark methods on the mouse data (per second)

روش‌های مورد مطالعه Study methods		وزن شش هفتگی W6W	شیب رشد Growth Slope	شاخص توده بدنی BMI	طول بدن Body Length
شبکه عصبی بیزی BRNN	NN=1	40.59 (0.68)	36.44 (1.49)	39.15 (1.07)	38.82 (0.58)
	NN=2	282.07 (0.59)	273.45 (4.61)	281.31 (0.51)	275.88 (0.28)
	NN=3	897.82 (1.89)	891.15 (7.06)	893.89 (0.99)	876.06 (1.27)
	NN=4	2047.66 (3.93)	2025.34 (5.28)	2036.63 (1.76)	1990.04 (2.47)
	NN=5	3888.58 (7.19)	3834.48 (6.93)	3859.87 (3.91)	3775.25 (5.6)
روش‌های پارامتری PM	GBLUP	348.26 (0.65)	347.67 (0.2)	350.17 (0.14)	348.22 (1.04)
	Bayes RR	131 (0.22)	131.03 (0.1)	131.77 (0.06)	131.61 (0.53)
	Bayes A	142.43 (0.1)	142.61 (0.08)	143.29 (0.06)	142.86 (0.43)
	Bayes B	160.43 (1.98)	151.67 (3.74)	148.53 (2.66)	150.69 (2.61)
	Bayes C	155.01 (0.42)	146.47 (1.2)	141.79 (1.06)	145.44 (1.23)
	Bayes L	147.75 (0.04)	147.95 (0.05)	148.72 (0.06)	148.38 (0.5)

BRNN: Bayesian Regularized Neural Networks
PM: Parametric Methods
NN: Number of neurons in hidden layer

نتیجه‌گیری کلی

باوجود بهتر بودن صحت و توانایی پیش‌بینی روش‌های پارامتری به دلیل معماری ژنتیکی افزایشی صفات موردبررسی، به نظر می‌رسد شبکه‌های عصبی بیزی ابزاری توانمند در پیش‌بینی ژنومی هستند که می‌توانند ارزش‌های اصلاحی ژنومی را با دقت مناسبی پیش‌بینی نمایند. توانایی پیش‌بینی ژنومی شبکه‌های عصبی به صفات هدف، گونه موردنظر و معماری شبکه عصبی بستگی دارد. درهرصورت قبل از استفاده از شبکه‌های عصبی بیزی در پیش‌بینی ژنومی بهتر است نتایج به‌دست‌آمده با روش‌های پارامتری مقایسه شود. همچنین قبل از استفاده عملی از شبکه‌های عصبی با تعداد نرون بیشتر در لایه مخفی در پیش‌بینی لازم است کارایی محاسباتی آنها بهبود یابد.

منابع

1- Calus, M., A. De Roos, and R. Veerkamp. 2008. Accuracy of genomic selection using different methods to define haplotypes. Genetics, 178(1), 553-561.

2- Chatterjee, N., B. Wheeler, J. Sampson, P. Hartge, S. J. Chanock, and J.H. Park. 2013. Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. Nature Genetics, 45(4), 400.

3- Clark, S. A., J. M. Hickey, and J. H. Van der Werf. 2011. Different models of genetic variation and their effect on genomic evaluation. Genetics Selection Evolution, 43(1), 18.

4- Colombani, C., P. Croiseau, S. Fritz, F. Guillaume, A. Legarra V. Ducrocq, and C. Robert-Granié. 2012. A comparison of partial least squares (PLS) and sparse PLS regressions in genomic selection in French dairy cattle. Journal of Dairy Science, 95(4), 2120-2131.

5- Crossa, J., P. Perez, J. Hickey, J. Burgueño, L. Ornella J. Cerón-Rojas, X. Zhang, S. Dreisigacker, R.Babu, Y. Li, and D. Bonnett. 2014. Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity, 112(1), 48.

6- Daetwyler, H. D., B. Villanueva, P. Bijma, and J. A. Woolliams. 2007. Inbreeding in genome-wide selection. Journal of Animal Breeding and Genetics, 124(6), 369-376.

7- Daetwyler, H. D., R. Pong-Wong, B. Villanueva, and J. A. Woolliams. 2010. The impact of genetic architecture on genome-wide evaluation methods. Genetics, 185: 1021-1031.

8- De los Campos, G., and P. Pérez-Rodríguez. "BGLR: Bayesian generalized linear regression." R package version 1.0.5 .2016.

9- De Los Campos, G., D. Gianola, G. J. Rosa, K. A. Weigel, and J. Crossa. 2010. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. Genetics Research, 92(4), 295-308.

10- Ehret, A., D. Hochstuhl, D. Gianola, and G. Thaller. 2015. Application of neural networks with back-propagation to genome-enabled prediction of complex traits in Holstein-Friesian and German Fleckvieh cattle. Genetics Selection Evolution, 47(1), 22.

11- Eichler, E. E., J. Flint, G. Gibson, A. Kong, S. M. Leal, J.H. Moore, and J.H.Nadeau. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. Nature Reviews Genetics, 11(6), 446.

12- Ghafouri-Kesbi, F., G. Rahimi-Mianji, M. Honarvar, and A. Nejati-Javaremi. 2017. Predictive ability of Random Forests, Boosting, Support Vector Machines and Genomic Best Linear Unbiased Prediction in different scenarios of genomic evaluation. Animal Production Science, 57(2), 229-236.

13- Gianola, D., and J. B. van Kaam. 2008. Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. Genetics, 178(4), 2289-2303.

14- Gianola, D., H. Okut, K. A. Weigel, and G. J. Rosa. 2011. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. BMC genetics, 12(1), 87.

15- Gianola, D., R. L. Fernando, and A. Stella. 2006. Genomic-assisted prediction of genetic value with semiparametric procedures. Genetics, 173: 1761-1776.

16- Goddard, M. 2009. Genomic selection: prediction of accuracy and maximisation of long-term response. Genetica, 136(2), 245-257.

17- Goddard, M. E., B. J. Hayes, and T. H. Meuwissen. 2010. Genomic selection in livestock populations. Genetics research, 92(5-6), 413-421.

18- González-Camacho, J., G. de Los Campos, P. Pérez, D. Gianola, J.E. Cairns, G. Mahuku, R. Babu, and J. Crossa, 2012. Genome-enabled prediction of genetic values using radial basis function neural networks. Theoretical and Applied Genetics, 125(4), 759-771.

19- Hayes, B. J., P. J. Bowman, A. C. Chamberlain, K. Verbyla, and M. E. Goddard. 2009. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genetics Selection Evolution, 41(1), 51.

20- Heslot, N., H. P. Yang, M. E. Sorrells, and J. L. Jannink. 2012. Genomic selection in plant breeding: a comparison of models. Crop Science, 52(1), 146-160.

21- Howard, R., A. L. Carriquiry, and W. D. Beavis. 2014. Parametric and nonparametric statistical methods for genomic selection of traits with additive and epistatic genetic architectures. G3: Genes, Genomes, Genetics, g3-114.
22- Lampinen, J., and A. Vehtari. 2001. Bayesian approach for neural networks—review and case studies. Neural networks, 14(3), 257-274.
23- Legarra, A., C. Robert-Granié, P. Croiseau, F. Guillaume, and S. Fritz. 2011. Improved Lasso for genomic selection. Genetics research, 93(1), 77-87.
24- Lin, Z., N. O. Cogan, L. W. Pembleton, G. C. Spangenberg, J. W. Forster, B.J. Hayes, and H.D. Daetwyler. 2016. Genetic gain and inbreeding from genomic selection in a simulated commercial breeding program for perennial ryegrass. The plant genome, 9(1).
25- MacKay, D. J., and D. J. Mac Kay. 2003. Information theory, inference and learning algorithms. Cambridge University Press.
26- Martini, J. W., N. Gao, D. F. Cardoso, V. Wimmer, M. Erbe, R.J. Cantet, and H. Simianer. 2017. Genomic prediction with epistasis models: on the marker-coding-dependent performance of the extended GBLUP and properties of the categorical epistasis model (CE). BMC Bioinformatics, 18(1), 3.
27- Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157(4), 1819-1829.
28- Meuwissen, T., and M. Goddard, 2010. The use of family relationships and linkage disequilibrium to impute phase and missing genotypes in up to whole-genome sequence density genotypic data. Genetics, 185: 1441-1449.
29- Mohammadi, Y., M. M. Shariati, S. Zerehdaran, M. Razmkabir, M. B. Sayyadnejad, and M. B. Zandi. 2016. The accuracy of genomic breeding value for production trait in Iranian Holstein Dairy Cattle using parametric and non-parametric methods. Animal Production, 18(1): 1-11 (In Persian).
30- Moser, G., S. H. Lee, B. J. Hayes, M. E. Goddard, N. R. Wray, and P.M.Visscher. 2015. Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. PLoS genetics, 11(4), e1004969.
31- Neves, H. H., R. Carvalheiro, A. M. P. O'brien, Y. T. Utsunomiya, A. S. Do Carmo, F.S. Schenkel, J. Sölkner, J.C. McEwan, C.P. Van Tassell, J.B. Cole, and M.V. Da Silva. 2014. Accuracy of genomic predictions in Bos indicus (Nellore) cattle. Genetics Selection Evolution, 46(1), 17.
32- Ogutu, J. O., H.-P. Piepho, and T. Schulz-Streeck. 2011. A comparison of random forests, boosting and support vector machines for genomic selection. In BMC proceedings (Vol. 5, No. 3, p. S11). BioMed Central.
33- Okut, H., X.-L. Wu, G. J. Rosa, S. Bauck, B. W. Woodward, R.D.Schnabel, J.F. Taylor, and D.Gianola. 2013. Predicting expected progeny difference for marbling score in Angus cattle using artificial neural networks and Bayesian regression models. Genetics Selection Evolution, 45(1), 34.
34- Okut, H., D. Gianola, G. J. Rosa, and K. A. Weigel. 2011. Prediction of body mass index in mice using dense molecular markers and a regularized neural network. Genetics Research, 93(3), 189-201.
35- Pérez, P., and G. de Los Campos. 2014. Genome-Wide Regression and Prediction with the BGLR Statistical Package. Genetics, 198(2), 483–495.
36- Pérez-Rodríguez, P., D. Gianola, J. M. González-Camacho, J. Crossa, Y. Manès, and S. Dreisigacker. 2012. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. G3: Genes, Genomes, Genetics, 2(12), 1595-1605.
37- Shaneh, A., and G. Butler. 2006. Bayesian learning for feed-forward neural network with application to proteomic data: the glycosylation sites detection of the epidermal growth factor-like proteins associated with cancer as a case study. In Conference of the Canadian Society for Computational Studies of Intelligence (pp. 110-121). Springer, Berlin, Heidelberg.
38- Tusell, L., P. Pérez-Rodríguez, S. Forni, X. L. Wu, and D. Gianola. 2013. Genome-enabled methods for predicting litter size in pigs: a comparison. Animal, 7(11), 1739-1749.
39- Wolc, A., J. Arango, P. Settar, J. E. Fulton, N. P. O'Sullivan, R. Preisinger, D. Habier, R. Fernando, D.J. Garrick, and J.C. Dekkers. 2011. Persistence of accuracy of genomic estimated breeding values over generations in layer chickens. Genetics Selection Evolution, 43(1), 23.
40- Xu, M., G. Zeng, X. Xu, G. Huang, R. Jiang, and W. Sun. 2006. Application of Bayesian regularized BP neural network model for trend analysis, acidity and chemical composition of precipitation in North Carolina. Water, Air, and Soil Pollution, 172(1-4), 167-184.
41- Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D.R. Nyholt, P.A. Madden, A.C. Heath, N.G. Martin, G.W. Montgomery, and M.E. Goddard. 2010. Common SNPs explain a large proportion of the heritability for human height. Nature Genetics, 42(7), 565.

Iranian Journal of Animal Science Research
Vol. 11, No. 3, Fall 2019, p. 377-388

نشریه پژوهشهای علوم دامی ایران
جلد 11، شماره 3، پاییز 1398، ص. 388-377

Genomic Enabled Prediction Using Bayesian Artificial Neural Networks and Parametric Methods a Comparative Study

S. Shadpour[1], M. Tahmoorespur[2]*, M. M. Shariati[3]
Received: 30-07-2018
Accepted: 05-01-2019

Introduction In genomic selection, genetic values of individuals are predicted using genetic markers that are distributed all across the genome and are in linkage disequilibrium with quantitative trait locus. Different methods have been introduced to predict genomic breeding values. These methods take into account different assumptions. Non-parametric methods, including artificial neural networks, have fewer assumptions than parametric methods, and can apply nonlinear relationships in genomic predictions so, in theory these approaches are more robust against genetic architecture changes and are able to provide better predictions.

Materials and Methods In current study, the prediction ability of Bayesian neural networks with different architectures (1 to 5 neurons in the hidden layer) and parametric methods (GBLUP, Bayes RR, Bayes A, Bayes B, Bayes C Bayes L) in four simulated genetic architectures and four real traits of mouse (six weeks weight, growth slope, body mass index and body length) were compared using the correlation coefficient between predicted and expected values, mean square error of prediction and computation time. All simulated genetic architectures were additive and the gene effects followed a normal distribution. The number of QTLs in the first and third genetic architecture was 50 and it was 500 for second and fourth genetic architecture. The heritability of the first and second genetic architectures was 0.3 and the heritability of the third and the fourth genetic architectures was 0.7. The real data consisted of 1,296 mice which were genotyped with 9,265 SNP markers.

Results and Discussion The highest prediction accuracy of Bayesian neural networks were 0.640 (4 neuron in the hidden layer), 0.664 (4 neuron in the hidden layer), 0.800 (1 neuron in the hidden layer) and 0.810 (1 neuron in the hidden layer), and the highest prediction accuracy of parametric methods were 0.711(Bayes B), 0.685 (Bayes A), 0.903(Bayes B) and 0.836 (Bayes B) respectively for one to four simulated genetic architectures. These results showed the superiority of parametric methods to Bayesian neural networks in terms of prediction accuracy in genetic architectures with additive effects. In additive genetic architectures, the allelic effects of genetic variations are independent. In parametric models, these effects are assumed to be independent, therefore in additive genetic architectures can be expected that parametric methods are able to provide better predictions than nonparametric methods.

The maximum predictive abilities of Bayesian neural networks to predict six weeks weight, growth slope, body mass index and body length were 0.474 (1 neuron in the hidden layer), 0.349 (4 neuron in the hidden layer), 0.154 (1 neuron in the hidden layer) and 0.214 (4 neuron in the hidden layer). The predictive abilities of parametric methods to predict these traits were similar and equal to 0.477, 0.336, 0.170, and 0.221 in average. The results showed that the predictive abilities of Bayesian neural networks and parametric methods were similar on real data as the difference between the best predictive ability of Bayesian neural networks and parametric methods for Six weeks weight, growth slope and body length were less than 1%. The difference was slightly higher for the body mass index and equal to 1.8%.

The mean squared error of prediction of Bayesian Neural Networks was slightly less than parametric methods in the simulated genetic architectures. The results indicate a slight superiority of Bayesian neural networks compared to parametric methods in terms of mean squared error of prediction as an indicator of overall fit. The mean square prediction error is an appropriate criterion for evaluating the prediction performance of different methods because it contains both accuracy and bias. Considering table (3) and table (5), it can be concluded that the prediction of the Bayesian neural network are less accurate but more unbiased than the parametric methods. This could be due to more applied penalty in parametric methods compared to Bayesian neural networks, which can lead to an increase in the average mean squared error of prediction. In real data, the mean squared error of prediction of the Bayesian neural networks and parametric methods were similar.

1,2,3- Ph.D., Professor and Assistant Professor of Animal Sciences Department, Faculty of Agriculture, Ferdowsi University of Mashhad, Iran.
(*- Corresponding Author Email: tahmoores@um.ac.ir)

The computation time of Bayesian neural networks was increased with an increase in the number of neurons in the hidden layer. The computation time of the parametric methods was the same with the exception of GBLUP. The GBLUP method took more computation time. The computation time of neural the networks with 1 to 2 neurons in the hidden layer were less than GBLUP. Genomic prediction using Bayesian Neural Networks with a greater number of neurons is really challenging, and improving their performance in terms of computational cost is necessary before applying them in genomic selection.

Conclusion Although parametric methods had better predictive accuracy and predictive ability due to the additive genetic architecture of the studied traits, it can be concluded that Bayesian neural networks are powerful tools in genomic enabled prediction that can predict genomic breeding values with acceptable accuracy.

The genomic prediction ability of the neural networks depends on target traits, the animal species, and neural network architecture. Before using Bayesian neural networks in genomic prediction, it is better to compare the results with parametric methods. It is also necessary to improve the computation time of the Bayesian neural networks with a greater number of neurons in hidden layer before applying them in real application of genomic selection.

Key words: Efficiency comparison, Genomic evaluation, neural networks, parametric methods.